

ОБЩАЯ ПСИХОЛОГИЯ, ПСИХОФИЗИОЛОГИЯ И ПСИХОФИЗИКА

В.М. Воронин

ВОСПРИЯТИЕ И ПОНИМАНИЕ ЕСТЕСТВЕННОЙ И СИНТЕЗИРОВАННОЙ РЕЧИ

Системы с речевым выходом, использующие синтезированную речь, все шире применяются в обучении, быту, промышленности и военной технике. Они применяются также в помощь лицам с нарушениями речи или зрения. Поэтому возникает необходимость в разработке надежных методов сравнения различных систем синтеза речи и методов оценки того, как человек воспринимает на слух и реагирует на речь, синтезируемую такими системами. Выбор системы с речевым выходом для конкретной практической задачи зависит от многих факторов, и разборчивость синтезированной речи является всего лишь одним из них. Нами был проведен сравнительный анализ результатов исследований зарубежных авторов, а также проведены собственные исследования, в которых для оценки восприятия синтезированной речи служили такие критерии, как разборчивость фонем, правильность распознавания слов и общее понимание смысла. Для того, чтобы сравнить звучание различных синтезаторов речи с естественной речью и больше узнать о восприятии речи, синтезируемой по алгоритмам, использовалось несколько методик. Полученные данные позволяют предположить, что такое восприятие зависит от взаимодействия нескольких факторов, в том числе от акустико-фонетических свойств речевого сигнала, условий восприятия и опыта слушателя. Различия в восприятии естественной речи и высококачественной синтезированной речи, по-видимому, связаны с избыточностью акустико-фонетической информации, закодированной в речевом сигнале.

Системы с речевым выходом в зависимости от способа синтеза речи можно разбить на два больших класса. К первому относятся системы речевого кодирования, использующие фиксированное число параметров для воспроизведения сравнительно ограниченного набора высказываний. Ценой потери гибкости с помощью этих систем можно получить разборчивую речь приемлемого качества (под гибкостью здесь понимается возможность широкого выбора среди различных высказываний). Ко второму классу относятся системы синтеза речи, генерируемой по алгоритмам. В отличие от систем первого класса, они дают менее разборчивую речь, которая звучит не вполне естественно. Однако такие системы обладают способностью преобразовывать в речь

любой текст, представленный в формате ASCII [1, 2]. За последнее время системы «текст – речь» (системы синтеза по тексту) были существенно улучшены за рубежом, поэтому во многом исчезли те преимущества, которые имели простые системы с кодируемым речевым выходом перед системами «текст – речь». При создании систем «текст – речь» проводятся интенсивные исследования, в основном сводящиеся к совершенствованию правил перевода букв в звуки, правил произношения фонем, а также методов синтеза полуслогов. Все это позволяет предположить, что в ближайшем будущем устройства с речевым ответом, работающие с произвольным текстом, смогут синтезировать разборчивую речь, звучащую достаточно качественно [9].

Отечественные системы синтеза речи появились в последние годы. Наиболее совершенной из них является Speaking Mouse Home, совместимая с операционной системой Windows 95 (Windows 98).

Поскольку качество речи, генерируемой системами «текст – речь», непрерывно улучшается, возникает необходимость в оценке и сравнении между собой характеристик различных синтезаторов речи, а также в систематической и надежной оценке устройств с речевым выходом. Это особенно заметно, если обратить внимание на характеристики сравнительно совершенных синтезаторов, таких как, например американские DECtalk, Prose-2000 или отечественной Speaking Mouse Home, которые обеспечивают синтез речи на разных языках и генерируют несколько разных синтетических голосов.

В этой связи важно знать, как на слух воспринимается и понимается синтезированная речь и насколько она отличается по качеству от естественной. Если бы у нас был набор алгоритмов или набор акустических критериев, с помощью которых можно было бы автоматически определять качество синтезированной речи, то не возникала бы проблема описания характеристик конкретной системы синтеза (или описания эффективности новых правил или методов синтеза). В этом случае можно было бы легко разработать соответствующие стандарты и повсеместно ими пользоваться.

К сожалению, пока еще нет метода автоматизации процесса оценки как естественной, так и синтезированной речи, поэтому их оценка производится на слух. Отсюда следует, что субъективно-статистические (аудиторские) испытания необходимо проводить в тех же условиях, в которых будет использоваться синтезированная речь.

Восприятие речи зависит как от слушателя, так и от свойств самого акустического сигнала, а также от набора алгоритмов, используемых при генерации такого сигнала [17]. Очевидно, что характеристики синтезаторов речи должны объективно оцениваться с применением

аудиторских испытаний. До недавнего времени лишь незначительное число работ зарубежных, в основном американских, авторов было посвящено исследованию разборчивости синтезированной речи (например, [7, 14, 18]). Что касается работ, где обсуждалась бы техническая сторона задачи определения разборчивости синтезированной речи, то их еще меньше (см. [16, 17]). Важно точно определить, какие именно параметры синтезированной речи нужно оценивать.

С одной стороны, восприятие и понимание синтезированной речи можно оценить с помощью комплекса объективных поведенческих экспериментов, которые дают точные и статистически надежные оценки характеристик системы с речевым выходом в конкретных условиях. С помощью этих экспериментов можно исследовать передачу слушателю лингвистической информации, содержащейся в речевом сигнале, и ответить на следующие вопросы: 1) насколько точно распознаются синтезированные фонемы и слова; 2) насколько правильно понимается синтезированное высказывание; 3) насколько легко воспринимается и понимается синтезированная речь. С другой стороны, возникает вопрос оценки приемлемости и естественности синтезированной речи, а также выяснения того, предпочитает ли слушатель тот или иной тип синтезатора. Вопрос о предпочтительности синтезированной речи нельзя решать, только применяя такие объективные критерии качества, как процент распознанных слов и время реакции. При ответе на него необходимо учитывать субъективные впечатления о качестве синтезированной речи.

Задаваемые при этом вопросы формулируются так, чтобы оценить степень естественности, приемлемости и предпочтительности. В лаборатории исследования речи университета штата Индиана на базе нескольких моделей «текст – речь» D.B. Piosonі было выполнено большое число исследований по восприятию синтезированной, то есть автоматически генерируемой по алгоритмам речи [5, 12, 13, 15, 16, 17, 18].

Цель исследований заключалась в том, чтобы получить фундаментальную информацию о восприятии синтезированной речи в хорошо контролируемых лабораторных условиях. Данные, полученные в этих исследованиях, могут послужить отправной точкой для последующего изучения человеческих факторов в практике применения систем с речевым выходом. В целом эти исследования были направлены на оценку правильности распознавания речи слушателями, которые должны были воспринимать и реагировать на синтезированную речь при самых разных требованиях и условиях эксперимента. В ходе исследований авторы провели также несколько сравнений правильности распознавания речи слушателями при решении стандартных задач слухового восприятия

с использованием синтезированной речи, генерируемой по алгоритмам на нескольких системах «текст – речь».

Мы считаем целесообразным привести некоторые результаты исследований американских авторов в тех аспектах, которые, на наш взгляд, важны как в целом для решения проблемы восприятия синтезированной речи, так и для конкретных человеко-машинных систем, использующих синтезированную речь.

1. Ограничения, влияющие на правильность распознавания

D.B. Piosoni и соавторы [15, 16, 17, 18] рассматривают ряд факторов, влияющих на правильность распознавания речи слушателем: 1) конкретные требования и условия, связанные со спецификой решаемой задачи; 2) ограничения, присущие системе обработки информации, которой наделен человек; 3) опыт и тренировка слушателя; 4) лингвистическая структура сообщения и 5) структура и качество речевого сигнала.

Разберем каждый из этих факторов подробнее.

1. Сложность задачи

Первым ограничивающим фактором является сложность задач, которые выполняет слушатель одновременно с восприятием речи. В ряде задач требования к реакции человека довольно просты, например решить, какое именно слово из двух заранее известных было произнесено. Другие задачи могут быть более сложными, например распознать некоторое неизвестное высказывание из фактически неограниченного набора, выполняя другую работу, которая тоже требует внимания. В литературе по когнитивной и инженерной психологии показано существенное влияние на человека, выполняющего некоторый круг задач по восприятию и распознаванию, таких факторов, как предлагаемый для восприятия набор, характер инструкций, субъективный ожидаемый результат, требования к вниманию [20]. На правильность распознавания оказывает также значительное влияние контекст и степень неопределенности задачи [8]. Поэтому, прежде чем делать какие-то определенные выводы относительно поведения человека или того, насколько правильно он распознает предъявляемый ему материал, необходимо понять все условия и требования конкретной задачи.

2. Ограничения, присущие человеку

Второй фактор, влияющий на распознавание синтезированной речи, связан со структурными ограничениями, присущими системе обработки информации, которой наделен человек, то есть с ограниченной способностью человека воспринимать, кодировать, запоминать и

извлекать из памяти информацию. Поскольку нервная система не может использовать сразу все параметры сенсорного раздражения (и поэтому вынуждена интегрировать звуковую энергию по времени), способность человека декодировать (расшифровывать) и запоминать исходные сенсорные данные, как оказалось, весьма ограничена. Для преодоления этих ограничений слушатель должен быстро преобразовать сенсорные входные сигналы в более абстрактную форму – в нейронные коды для надежного хранения в памяти и последующей обработки. В результате проведенных исследований по когнитивным и перцептивным процессам выяснилось, что главное ограничение при обработке информации налагает кратковременная память человека (КВП) [19]. Количество информации, которое может обрабатываться в кратковременной памяти и передаваться на следующий уровень, в значительной мере зависит от степени внимания слушателя, его прошлого опыта, а также от «качества» входной сенсорной информации.

3. Опыт и тренировка

Третий фактор связан со способностью оператора (слушателя) быстро обучаться эффективным стратегиям восприятия и опознания для улучшения правильности распознавания почти в любой задаче. Если человеку обеспечить условия для обучения при наличии соответствующей обратной связи, то он может научиться классифицировать новые стимулы, запоминать сложные последовательности входных сигналов и реагировать на быстрые изменения входных сигналов в разных сенсорных модальностях. Очевидно, что способность слушателя гибко адаптироваться к требованиям конкретной задачи является важным фактором, который вносит определенные ограничения и который следует учитывать и управлять им в процессе оценки восприятия синтезированной речи оператором.

4. Множество сообщений

Четвертый фактор связан со структурой множества (набора) сообщений. Иными словами, этот фактор связан с ограничениями на число возможных сообщений, а также на их организацию и лингвистические свойства. Набор сообщений может состоять из слов, которые различаются только одной фонемой, или из слов и фраз, которые сильно отличаются по длине, характеру ударения и фонотактическим структурам. Использование этих особенностей слушателями зависит от их лингвистических познаний [11]. Выбор определенных звуков речи и образование из них слов ограничены и подчиняются фонологическим правилам языка, формирование предложений из отдельных слов подчиняется правилам синтаксиса и, наконец, смысл отдельных слов и общий смысл предложений в тексте определяется семантикой и

прагматикой языка. Вклад этих уровней лингвистических структур в восприятие сильно меняется в зависимости от того, что воспринимается – изолированные слова, предложения или отрывки беглой речи.

5. Характеристики сигнала

Пятый фактор связан с акустико-фонетической и просодической структурой, то есть со сходством звучания синтезированного речевого сигнала с естественной речью. Речевые сигналы можно рассматривать как физическую реализацию сложной иерархически организованной системы лингвистических правил, с помощью которых свойства речевого сигнала ограничиваются акустикой голосового тракта. Акустико-фонетическая структура естественной речи отражает эти физические ограничения. Синтезированные сигналы являются упрощенными сигналами, звучание которых определяется лишь ограниченным подмножеством того множества акустических параметров, которые используются для передачи фонетической информации в естественной речи. Кроме того, в сравнении с естественной речью, акустические параметры, используемые для представления текста в синтезированной речи, значительно стилизованы и не могут передать фонетический контекст.

II. Аудиторская оценка естественной и синтезированной речи

Ошибки в системе «текст – речь», которые влияют на общую разборчивость речи, могут происходить в основном в трех случаях: 1) при переводе букв и буквосочетаний в звуки; 2) при вычислении и воспроизведении супraseгментной структуры; 3) в правилах фонетического исполнения, когда внутреннее представление фонем и слогов преобразуется в звуковые колебания.

1. Разборчивость фонем

В работах американских авторов по оценке синтезированной речи чаще всего использовался метод, называемый модифицированным райм-тестом [4, 14]. Сейчас этот тест принят в качестве основного для оценки сегментной разборчивости синтезированной речи.

При модифицированном райм-тесте (MPT) слушатель должен идентифицировать одиночные слова, делая выбор при ответе из предлагаемых ему шести слов, которые отличаются друг от друга одной буквой в начале либо в конце слова [6]. Примером являются односложные слова типа «согласная – гласная – согласная» (СГС): МИР – ТИР – ЛИР; РИС – РИМ; ДОМ – КОМ; МОР – СОР; СОР – СОК. MPT позволяет оценить эффективность идентификации слушателем либо начальной, либо конечной фонемы в наборе произнесенных слов. В одном из исследований было проведено сопоставление естественной речи с синтезированной

речью, создаваемой пятью различными системами «текст – речь», такими как Type-³N-Talk, Speech Plus Prose-2000, экспериментальной системой MITalk-79, Infolox и DECTalk [5]. Основные результаты применения MPT – наиболее высокая правильность узнавания (99,4% правильных ответов) получена при естественной речи. В случае системы DECTalk здесь оценивалась речь, воспроизводимая двумя (из девяти) голосами, имеющими наименования «Поль» (Paul) и «Бетти» (Betty). Для этих голосов точность идентификации оказалась разной – правильно было идентифицировано 96,7% слов, произносимых голосом «Поль», и только 94,4% слов, произносимых голосом «Бетти». Точность идентификации голоса «Поль» наиболее приближена к точности идентификации естественной речи и заметно превышает соответствующий показатель для всех других исследованных данными авторами систем «текст – речь». Правильность распознавания речи, синтезированной системой MITalk, была несколько ниже, чем для двух из голосов системы DECTalk (93,1% правильно идентифицированных слов). Синтезированная речь, создаваемая опытным образцом системы Prose-2000, правильно идентифицировалась в 87,6% случаев. На усовершенствованном варианте 3.0 данной системы удалось достигнуть правильности распознавания – 94,3%. Для многоязычной системы Infovox синтезированная английская речь была правильно идентифицирована в 87,4% случаев. Votrax Type-Talk давала наименее разборчивую синтезированную речь, правильность распознавания которой составила всего лишь 67,2%. Эти результаты, полученные в одних и тех же лабораторных условиях при строгом соблюдении неизменности условий эксперимента, показали, что системы «текст – речь» сильно отличаются друг от друга. На наш взгляд, эти различия напрямую связаны с глубиной фундаментальных исследований, проводившихся с целью разработки правил фонетического исполнения для этих систем с речевым ответом.

D.B. Piosoni и сотрудники помимо стандартной процедуры MPT с ограниченным набором вариантов, из которых производится выбор, провели также испытания на основе MPT при неограниченном выборе. Согласно этой процедуре, в каждой экспериментальной попытке слушатель должен был записать то слово, которое он услышал. Данная методика позволяет оценить правильность распознавания в условиях, когда ограничения, налагаемые на возможные ответы, минимальны (ответом может быть любое известное слушателю слово типа «согласная – гласная – согласная», а не одно из шести предлагаемых слов, как в стандартной процедуре MPT). Данная методика позволяет получить также сведения относительно разборчивости гласных, что невозможно в методике MPT с выбором из ограниченного числа возможных ответов.

Сравнение восприятия речи, синтезированной различными системами «текст – речь», в двух вариантах МРТ (с ограниченным и неограниченным выбором), и восприятия естественной речи позволяет оценить, насколько слушатель опирается на ограничения, налагаемые на набор ответов.

При выполнении текста МРТ с неограниченным выбором число правильно идентифицированных слов естественной речи составило 97,2%, тогда как при ограниченном выборе – 99,4%. Правильность распознавания естественной речи даже при отсутствии ограничений на набор ответов была выше, чем речи, синтезированной с помощью любой системы «текст – речь» при ограниченном числе возможных ответов. При МРТ с неограниченным выбором правильность распознавания слов, синтезированных на системе MITalk-79, оказалась значительно ниже – 75,4%. Соответствующий показатель для речи, воспроизводимой голосом «Поль» на DECTalk – 86,7%, а голосом «Бетти» - 82,5%. Эти результаты указывают на наличие сильной взаимосвязи между разборчивостью речи при выполнении МРТ в двух вариантах – с ограниченным и неограниченным выбором. Несмотря на то, что порядок разборчивости для этих двух вариантов МРТ остается одинаковым, не вызывает сомнений, что по мере того, как речь становится менее разборчивой, слушатели, пытаясь распознать слова, все больше полагаются на ограничения, наложенные на возможные варианты ответа.

Мы провели аналогичные исследования с применением модифицированного райм-теста с использованием двух синтезаторов отечественных лабораторий.

Первый синтезатор осуществлял упрощенный синтез, работал с операционной системой MS DOS, а второй – Speaking Mouse Home – являлся двуязычным (русский, английский), десятиголосовым синтезатором. В наших экспериментах при ограниченном наборе слов (шесть слов в наборе) естественная речь правильно была опознана в 99,8% случаев предъявления. Синтезированная речь при использовании упрощенного синтезатора была правильно опознана в 75% случаев, а при использовании синтезатора Speaking Mouse Home голосом «Диктор» была правильно идентифицирована в 98,5% случаев предъявления. При использовании МРТ с неограниченным набором слов показатели выглядят следующими образом:

естественная речь – 98,3%

упрощенный синтезатор – 67,2%

Speaking Mouse Home (голос «Диктор») – 95,3%

Speaking Mouse Home (голос «Mouse») – 89,2%

Speaking Mouse Home (голос «King Kong») – 87,7%

Таким образом, можно констатировать, что идентификация слов синтезированной речи, генерируемой системой Speaking Mouse Home, практически приближается к распознаванию слов естественной речи и при произношении голосом «Диктор» выше, чем результаты, полученные американскими авторами.

2. Распознавание слов в предложениях

Теперь интересно сравнить результаты, полученные американскими авторами, с нашими результатами при распознавании слов в предложениях. Для того, чтобы оценить влияние лингвистических закономерностей на правильность восприятия речи, американские авторы провели сравнение правильности распознавания слов в двух типах предложений [5]. Предложения первого типа были осмысленными и синтаксически корректными (гарвардские психоакустические предложения [3]). Пример такого предложения:

1) Add salt before you fry the egg. (Посолите яичницу заранее).

Предложения второго типа были корректными синтаксически, но неправильными семантически (синтаксические предложения Хаскина [14]). Такие испытательные предложения синтаксически построены как обычные предложения, но не имеют смысла. Пример такого предложения:

2) The old farm cost the blood. (Старая ферма стоит крови).

Сравнив правильность распознавания слов в двух этих типах предложений, можно оценить влияние, которое оказывает на распознавание слов смысл предложения и лингвистические закономерности [5]. В табл. 1 указан процент правильно идентифицированных слов для осмысленных и семантически бессмысленных предложений в случае как естественной, так и синтезированной речи (система MITalk-79, опытный образец Speech Plus Prose-2000 и DECTalk, голоса «Поль» и «Бетти»).

Таблица 1

Процент правильно опознанных слов в
осмысленных и семантически бессмысленных
предложениях в экспериментах американских авторов

Речь	Тип предложения	
	Осмысленное в %	Бессмысленное в %
Естественная	99,2	97,7
MITalk-79	93,3	79,7
DEC Paul	95,3	86,8
DEC Betty	90,5	75,1

Результаты наших экспериментальных исследований с двумя типами речевых синтезаторов и дикторской (естественной) речью отражены в табл. 2.

Таблица 2

Процент правильно опознанных слов в осмысленных и семантически бессмысленных предложениях в наших экспериментах

Речь	Тип предложения	
	Осмысленное в %	Бессмысленное в %
Естественная	99,5	98,2
Упрощенный синтезатор	85,2	73,1
Speaking Mouse Home (голос «Диктор»)	97,3	86,2

Эти результаты свидетельствуют, что как для естественной, так и для синтезированной речи распознавание слов в осмысленных предложениях было более уверенным. Кроме того, сравнение правильности идентификации слов в предложениях двух указанных типов выявило, что, по мере того, как речь становится все менее разборчивой, слушатели в гораздо большей степени начинают опираться на семантические закономерности.

3. Понимание речи на слух

Понимание устной, в том числе разговорной речи, является весьма сложным познавательным процессом, который включает в себя расшифровку сенсорной информации, поиск ранее запомненной информации в долговременной памяти и последующую интерпретацию и интеграцию различных источников знаний, которыми обладает слушатель [10, 15]. Поэтому понимание речи зависит от большого числа разнообразных и сложных факторов, многие из которых к настоящему времени еще недостаточно изучены в когнитивной психологии. Из-за взаимодействия нескольких разных источников знаний, участвующих в процессе понимания, количественно оценить этот процесс трудно. Эта трудность усугубляется еще и тем, что отсутствует соответствующая теоретическая модель речи, которая могла бы помочь при разработке соответствующих процедур измерения. Более того, нет и таких теоретических моделей, которые бы учитывали различные стратегии,

используемые слушателями для облегчения понимания речи при разных условиях прослушивания и при разных требованиях, налагаемых задачах.

Одним из факторов, который, судя по всему, играет важную роль в понимании речи на слух, является качество исходного сигнала, то есть разборчивость речи. Однако акустико-фонетические свойства речевого сигнала – это только один из источников информации, которым пользуется человек при восприятии и понимании разговорной речи. Как видно из приведенных выше результатов, дополнительно нужно учитывать вклад более высоких уровней лексических знаний в восприятие и понимание.

Вначале, пытаясь количественно оценить понимание синтезированной речи, мы хотели определить, насколько хорошо слушатели понимают лингвистическое содержание беглой речи, синтезируемой системой «текст – речь» Speaking Mouse Home. По нашему мнению, проблемам, связанным с восприятием лингвистического содержания как естественной, так и синтезированной (в частности, отрывков осмысленной беглой) речи, действительно, уделялось мало внимания.

Чтобы оценить понимание речи, мы выбрали несколько стандартных тестов (отрывков повествовательного характера) для оценки понимания прочитанного взрослыми испытуемыми, а также соответствующий набор тестовых вопросов с несколькими альтернативами. Отрывки были разными, охватывали широкий круг тем и отличались по стилю изложения и по словарному запасу.

При выборе отрывков учитывался также интерес для слушателей, участвовавших в лабораторных экспериментах по оценке речи. Поскольку отрывки выбирались из текстов для чтения разных типов, они были различны по трудности восприятия и стилю. Эти различия позволяли оценивать вклад всех отдельных компонентов системы «текст – речь» в понимание речи с единой позиции. Мы полагаем, что результаты этих экспериментов послужат основой, с помощью которой можно оценивать систему «текст – речь» в целом на материалах, сравнимых с теми, которые используются в приложениях, например, в читающих машинах или системах поиска информации в базе данных.

В нашем первом исследовании участвовали три группы испытуемых (они впервые принимали участие в подобных экспериментах) по 15 человек в каждой группе. Первая группа прослушивала отрывки текстов, выводимые с помощью системы Speaking Mouse Home, вторая группа – естественную речь, а испытуемые, входящие в третью группу, читали про себя эти же тексты, используя визуальное восприятие. После

каждого отрывка текста все три группы слушателей отвечали на один и тот же набор вопросов.

Результаты экспериментов по пониманию для всех групп выглядят следующим образом. При усреднении данных по последним 10 отрывкам текста оказалось, что степень понимания прочитанного текста выше (на 9%, $p < 0,01$), чем тогда, когда текст воспроизводился системой Speaking Mouse Home. Однако такое различие между результатами для разных групп наблюдалось, в основном, в первой половине испытаний. Во второй половине степень понимания для групп, слушавших синтезированную речь, заметно повысилась, тогда как степень понимания текста группой слушателей, которые сами читали текст, практически не изменилась. Хотя в целом степень понимания текста группой слушателей, в работе с которыми использовалась естественная речь, была несколько выше, чем при прослушивании синтезированной речи, в этой группе улучшения во второй половине испытаний практически не наблюдалось.

Улучшение степени понимания во второй половине испытаний для слушателей, которые прослушивали синтезированную речь, согласуется с ранее полученным результатом по распознаванию слов в предложениях, воспроизводимых синтезированным голосом, когда уже после короткого периода прослушивания правильность распознавания повышается. Это позволяет предположить, что в целом различие между группами по степени понимания текста объясняется, по-видимому, степенью привыкания к синтезированной речи и не связано с какими-либо тонкими различиями в основных приемах, которые используют слушатели при восприятии и понимании содержания этих отрывков текста.

Выводы

Результаты применения модифицированного райм-теста выявили довольно высокий уровень сегментной разборчивости речи, синтезированной на системе Speaking Mouse Home. Для системы упрощенного синтеза результаты таких испытаний оказались гораздо хуже. Постоянное совершенствование приводит к синтезаторам все более высокого класса. Можно ожидать, что при условии дополнительных научно-исследовательских и опытно-конструкторских разработок речь, синтезируемая высококачественными системами «текст – речь», по разборчивости вскоре приблизится к естественной речи в условиях лабораторных экспериментов.

Результаты экспериментов с двумя типами предложений показали, что контекст является важным средством, облегчающим распознавание. Когда слушатели имеют как семантическую, так и синтаксическую информацию, правильность распознавания слов в предложениях

повышается. Однако если использовать семантические знания затруднительно или даже вообще невозможно, слушатели должны в силу необходимости полагаться главным образом на акустико-фонетическую информацию, которую несет сигнал, а также на свои знания в области морфологии. Очевидно, что более высокая степень правильности распознавания предложений обусловлена источниками знаний высших уровней, а в отсутствие таких знаний правильность распознавания заметно хуже.

Наконец, результаты проведенных экспериментов по пониманию на слух показывают, что слушатели в состоянии правильно отвечать на вопросы с несколькими альтернативными ответами, касающиеся содержания отрывков, воспроизводимых с помощью синтезатора беглой речи. Уже через несколько минут прослушивания синтезатора речи понимание заметно улучшается и в конце концов приближается к той степени, которая отмечается, когда слушатель сам читает данный текст или прослушивает его в дикторском исполнении (естественная речь). Имеется, однако, ряд трудностей при определении степени понимания тех материалов, которыми мы пользовались. Во-первых, материалы предназначались для определения степени понимания при чтении, а не при слушании. Поэтому предполагалось, что при испытаниях читатель имеет возможность перечитать материал, чтобы ответить на некоторые вопросы. Читатель всегда имеет доступ к любому месту текста, тогда как слушатель не может снова прослушать уже прозвучавшие части текста. Во-вторых, методика многоальтернативных вопросов не позволяет непосредственно оценивать процессы восприятия, используемые при расшифровке входного речевого сигнала. Более того, эти вопросы позволяют оценивать степень понимания уже после того, как текст был предъявлен. Поэтому в данной ситуации нужно учитывать влияние постперцептивных стратегий понимания и индивидуальной установки слушателя. Таким образом, многоальтернативные вопросы не позволяют количественно оценить реальновременные когнитивные механизмы, участвующие в понимании, а просто показывают конечный результат понимания. Поэтому наша дальнейшая работа будет направлена на выявление этих механизмов, а также на исследования в области практического применения систем с речевым выходом.

СПИСОК ЛИТЕРАТУРЫ

1. *Allen J.* Reading machines to the blind: The technical problems and methods adopted for their solution, *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 259-264, 1973.

2. *Allen J.* Linguistic-based algorithms offer practical text-to-speech system // *Speech Technol*, vol. 1, no. 1, pp. 12-15, 1981.
3. *Egan J.P.* Articulation testing methods // *Laryngoscope*, vol. 58, pp. 955-911, 1948.
4. *Fairbanks G.* Test of fonemic differentiation: The rhyme test // *J. Acoust. Soc. Amer.*, vol. 30, pp. 596-600, 1958.
5. *Greene B.G., Manous L.M., Piosoni D.B.* Preliminary evolution of DECtalk // *Speech Res. Lab. Tech. Note* 84-03, Bloomington, IN, Indiana University, 1984.
6. *House A.S., Williams C.E., Hesker M.H., Kryter K.* Articulation-testing methods: Consonantal differentiation with a closed – response set // *J. Acoust. Soc. Amer.* vol. 37, pp. 158-166, 1965.
7. *Ingemann F.* Speech synthesis by rule using the FOVE program // *Haskins Lab. Status Rep. On Speech Research* SR-54, pp. 165-173, 1978.
8. *Kantowitz B.H., Sorkin R.D.* Human Factors: Understanding People-Systems Relationships. New York. Wiley, 1983.
9. *Klatt D.N.* Timing rules in Klattalk: Implications for models of speech production // *J. Acoust. Soc. Amer.*, Suppl. 1, vol. 73, pp. 566, 1983.
10. *Marslen-Wilson W.* Function and process in spoken word recognition – A tutorial review // *H. Bouma and D.G. Bouwhuis, Eds., Attention and Performance* vol. X. Hillsdale, Nj: Earbaum, 1964.
11. *Miller G.A., Hiese G., Lichten W.* The intelligibility of speech as a function of the conext of materials” // *J. Experimental Psychol*, vol. 41, pp. 329-335, 1951.
12. *Nusbaum H.C., Piosoni D.B.* Some constraints on the perception of synthetic speech, *Behavior Res. Methods instrum.* 1983.
13. *Nusbaum H.C., Piosoni D.B.* Perceptual evaluation of synthetic speech: Some constraints on the use of voice response systems // *Proc. 3rd Voice Data Entry Systems Applications Conf.* Sunnyvale, CA, Lockheed, 1983.
14. *Nye P.W., Gaitenby J.* The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences // *Haskins Lab. Status Rep. On Speech Res.*, vol. 38, pp. 169-190, 1974.
15. *Piosoni D.B.* Speech perception // *W.K. Estes, Ed., Handbook of Learning and Cognitive Processes: Volume 6*, Hillsdale, Nj.: Erlbaum, 1978.
16. *Piosoni D.B.* Some measures of intelligibility and comprehension // *Allen, S. Hunnicutt, and D.H. Klatt, Eds., Conversion of Unrestricted Text to Speech*, Notes for MIT Summer Couse 6.69s, July 1979.
17. *Piosoni D.B.* Perception of speech: The human Listener as a cognitive interface // *Speech Tecnol.*, pp. 10-23, 1982.

18. Piosoni D.B., Hunnicutt A.S. Perceptual evaluation of MITalk: The MIT unrestricted text-to-speech system // 1980 IEEE int. Cont. Rec. On Acoustics, Speech, and Signal Processing, pp. 572-575, 1980.
19. Shiffrin R.M. Capacity limitations in information processing attention, and memory // W. K. Estes, Ed., Handbook of Learning and Cognitive Processes, vol. 4, Hillsdale, Nj: Erlbaum, 1976.
20. Wickens C.D. The structure of attentional resources // R.S. Nickerson, Ed., Attention and Performance VIII. Hillsdale, Nj: Erlbaum, 1984.

О.Е. Сурнина, Н.В. Антонова

ОСОБЕННОСТИ ВОСПРИЯТИЯ ВРЕМЕНИ ЛЮДЬМИ ПОЖИЛОГО ВОЗРАСТА

В работе исследуются особенности отражения времени пожилыми людьми. В опытах принимали участие 27 мужчин и 27 женщин в возрасте от 60 до 91 года. В качестве контрольной группы – 50 молодых испытуемых в возрасте 17 – 25 лет. Они отмеривали временные интервалы 1, 3, 5, 7 и 10 секунд. Было обнаружено, что пожилые люди в значительной степени переотмеривают 1-секундный интервал, остальные – недоотмеривают. Наибольшая ошибка измерения отмечается у женщин. Субъективная временная шкала у пожилых испытуемых уже физической, а показатель степени психофизической функции отмеривания достоверно меньше единицы (при $p \leq 0,05$). Молодые испытуемые недоотмеривают все длительности данного диапазона, а субъективная шкала времени близка к физической.

Отражение времени является неотъемлемой составной частью целостного мировосприятия. Деятельность человека с самого рождения организована во времени и от корректности этой организации зависит уровень адаптации человека к окружающей среде.

Как отмечается в литературе, адекватность восприятия времени зависит от целого ряда факторов, ведущим из которых является возраст [1, 2, 21]. В абсолютном большинстве случаев в исследованиях принимают участие испытуемые, возраст которых находится в пределах от 7 до 30 лет, реже исследуется дошкольный период [12, 14, 19]. Как у нас в стране, так и за рубежом такие исследования хотя и проводятся, но носят несистематический характер [16, 21]. Вместе с тем, учитывая хорошо известный факт, что люди пожилого возраста составляют значительную часть населения нашей страны, представляется достаточно важным изучение особенностей их восприятия окружающего мира вообще и восприятия времени в частности. Здесь возникает целый ряд вопросов.